Tracking on GPU at LHCb's full software trigger

Alessandro Scarabotto on behalf of the LHCb Collaboration

LPNHE (Paris)

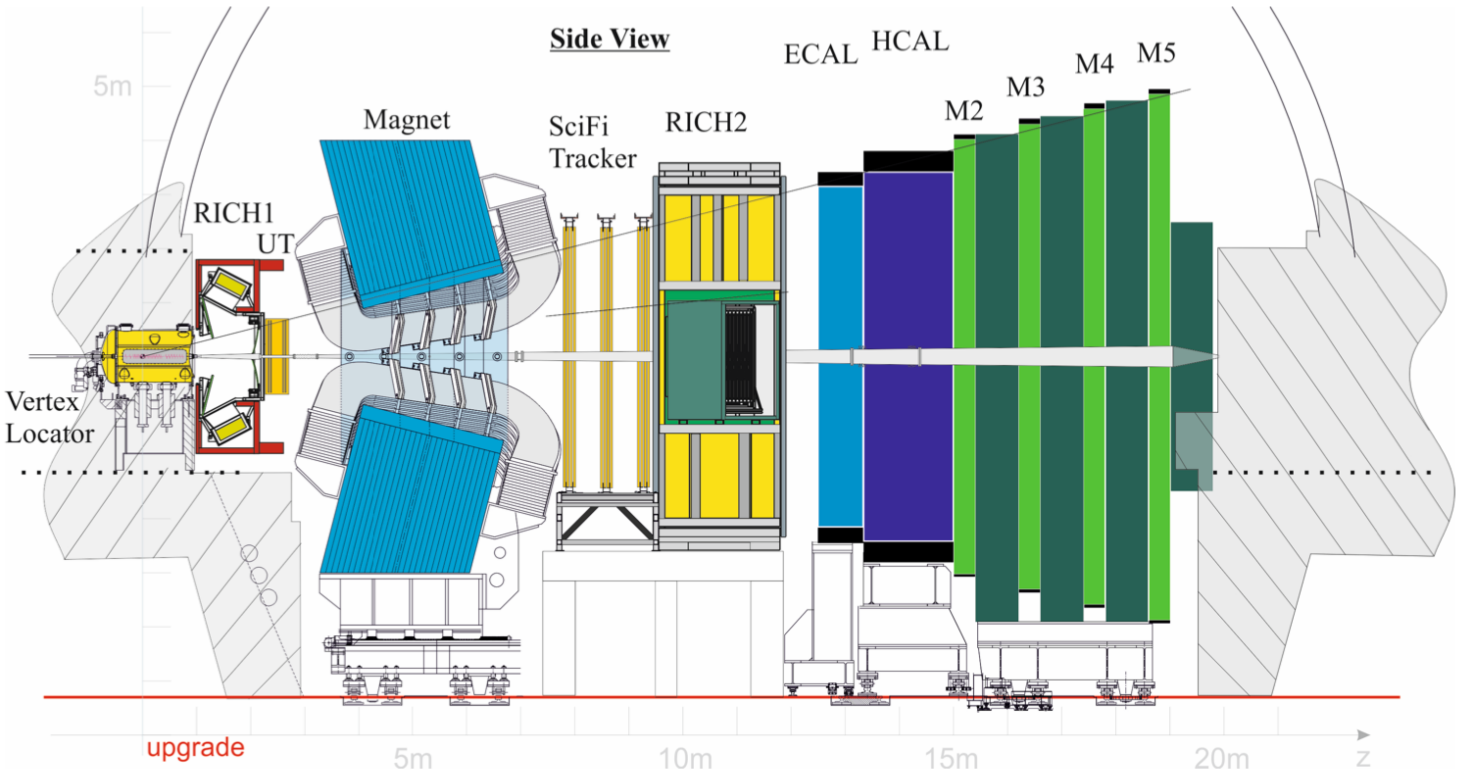
Connecting The Dots 2022

Princeton, USA

May 31 – June 2, 2022

LHCb detector Upgrade

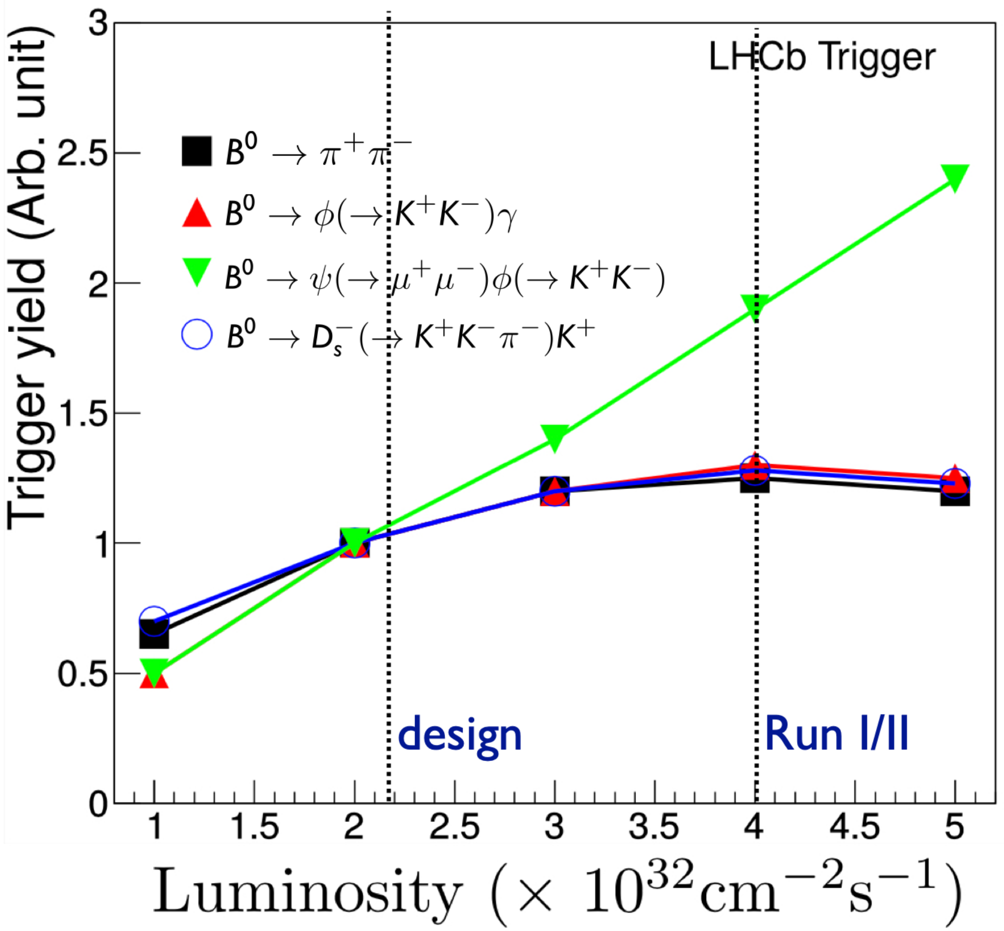
- Major upgrade to record $15\ fb^{-1}$ in Run3
- Luminosity increased by a factor 5:
  - $\mathcal{L} = 2 \cdot 10^{33}\, cm^{-2}s^{-1}$
- Upgrade of most subdetectors and readout system



LHCb-TDR-12

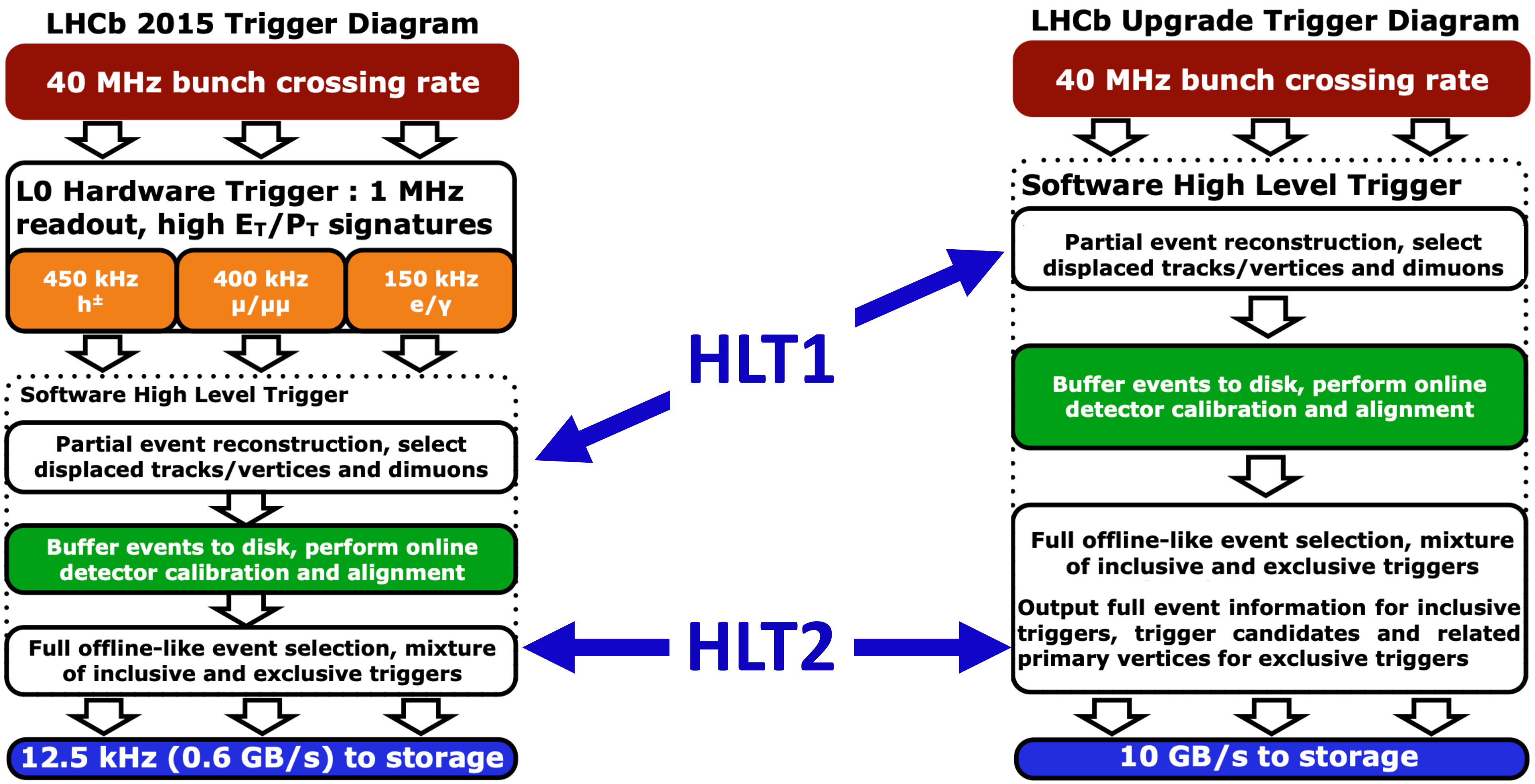
Full software trigger

- Hardware trigger removed
- Saturation of many hadronic channels already at Run2 luminosity
- Full reconstruction of events at LHC rate of 40 MHz by a software-based trigger
- Increase of hadronic channels' triggering efficiency by a factor 2 to 4 wrt Run2



J. Phys.: Conf. Ser. **878** 012012

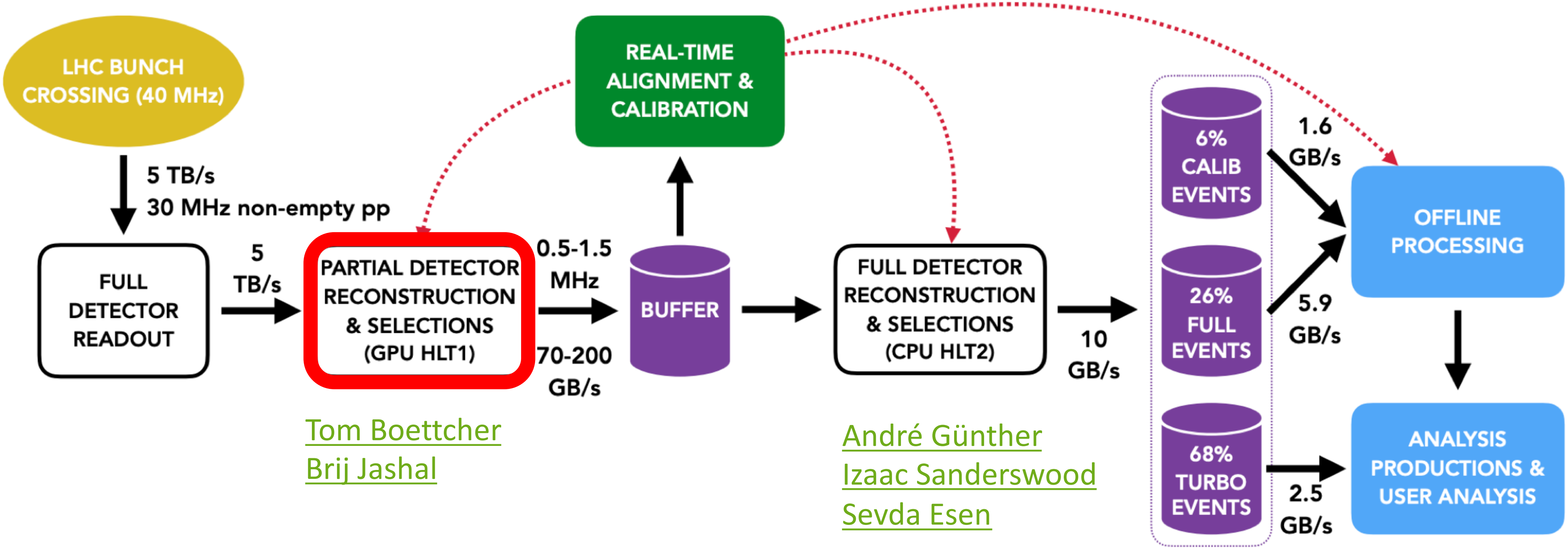
Trigger Upgrade



LHCb data flow

- HLT1 software reconstruction fully on GPUs at non-empty pp collisions of 30 MHz rate

Florian Reiss



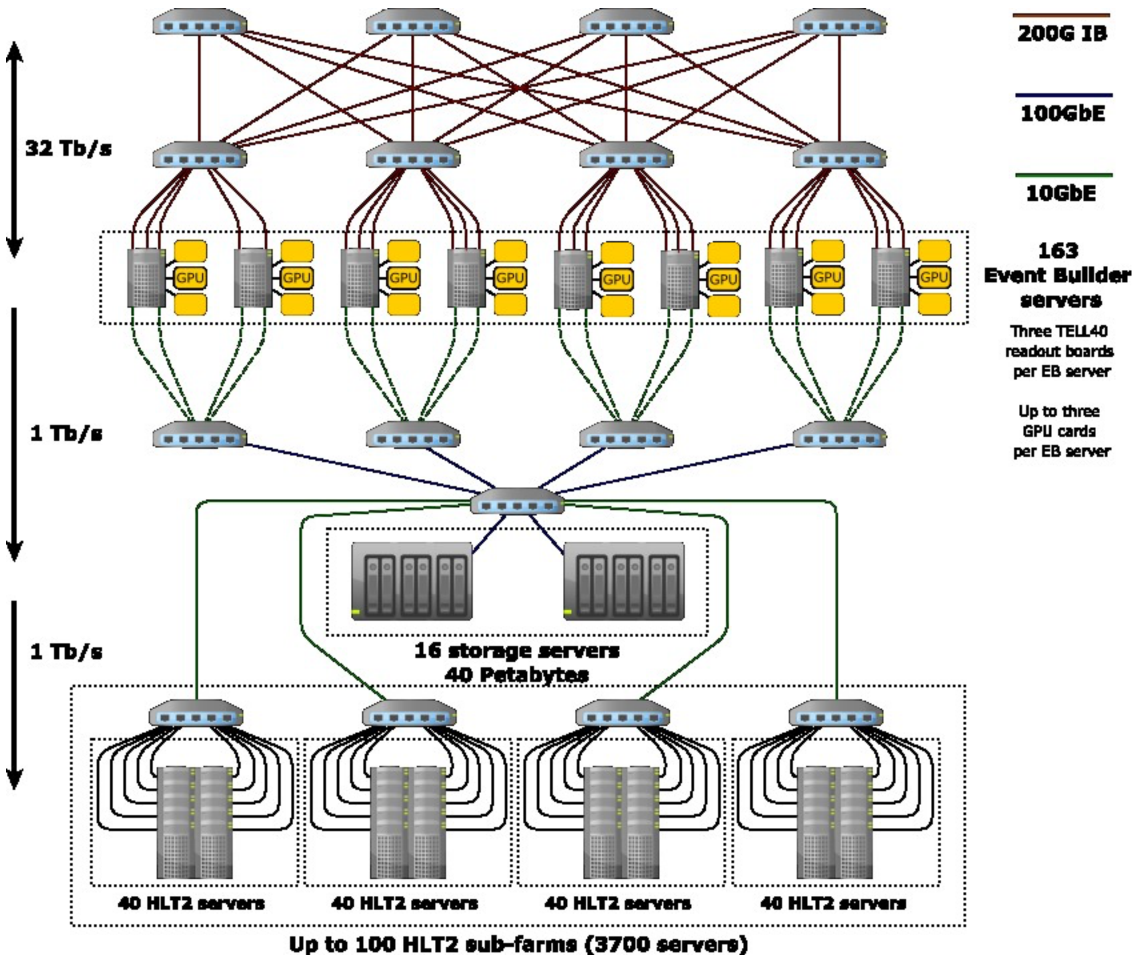
Tom Boettcher
Brij Jashal

André Günther
Izaac Sanderswood
Sevda Esen

LHCB-FIGURE-2020-016

Why using GPUs?

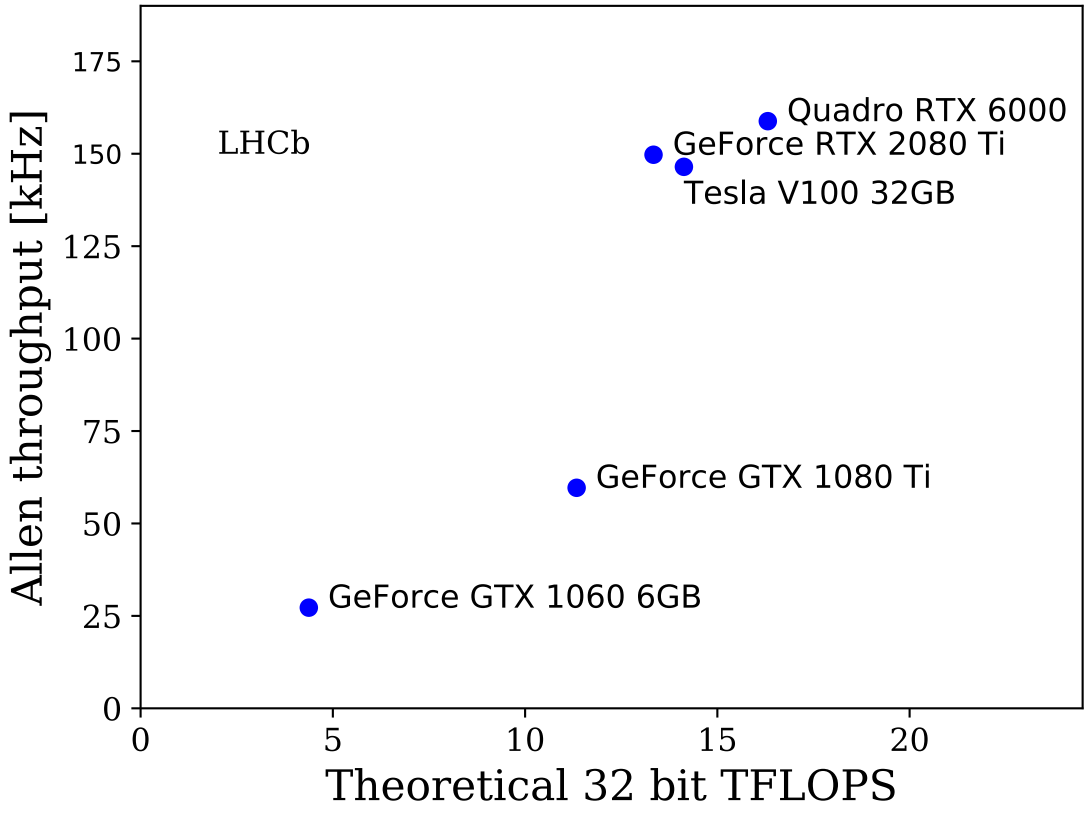
- Raw detectors info sent to a custom data processing centre
- FPGA cards receive data at average 32 Tb/s
- Event Builder (EB) CPU servers produce events packets
- EB servers contain slots to be used for up to 3 GPUs (zero overhead costs)
- Most of the HLT1 tasks naturally lend themselves to a very high degree of parallelism



[Comput.Softw.Big Sci. 6 (2022) 1, 1]

Why using GPUs?

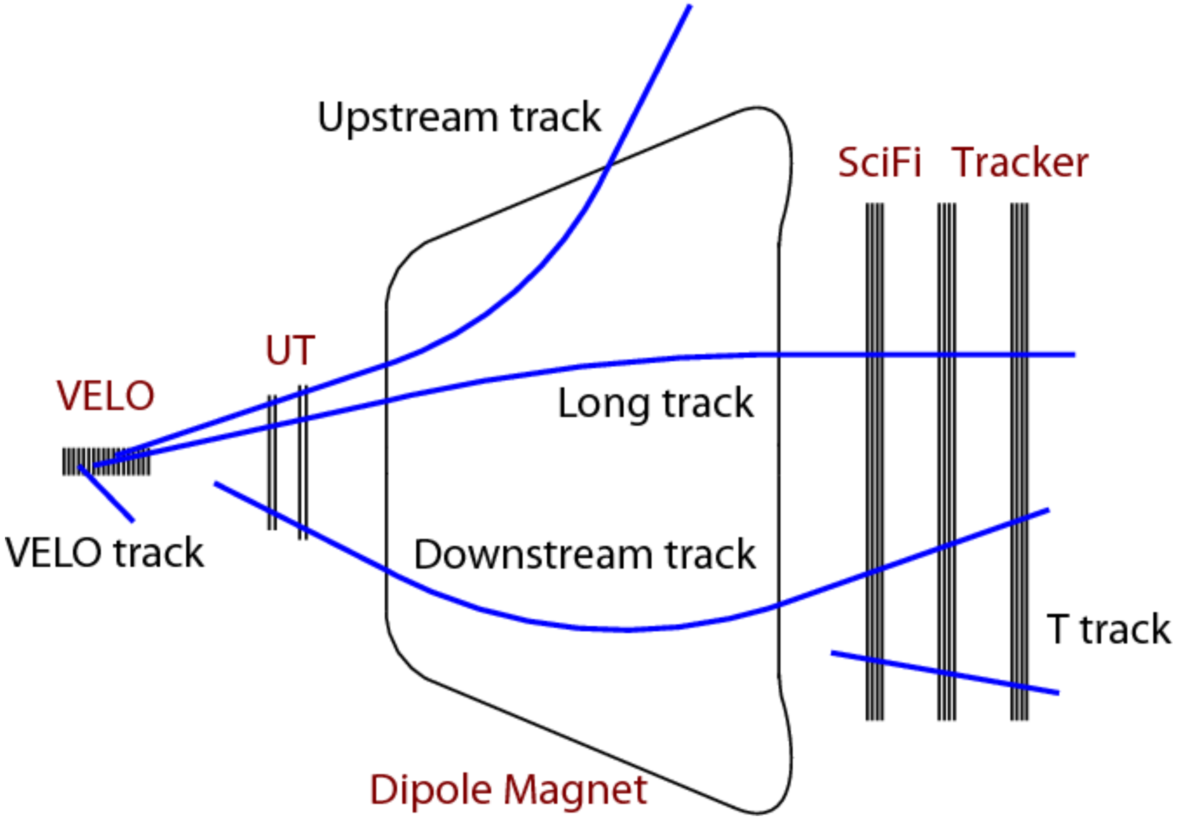
- Allen project
- Throughput scales linearly as a function of the theoretical peak 32-bit FLOPS performance
- Higher luminosity can be handled by improved GPU technology (no saturation expected)



LHCB-FIGURE-2020-014

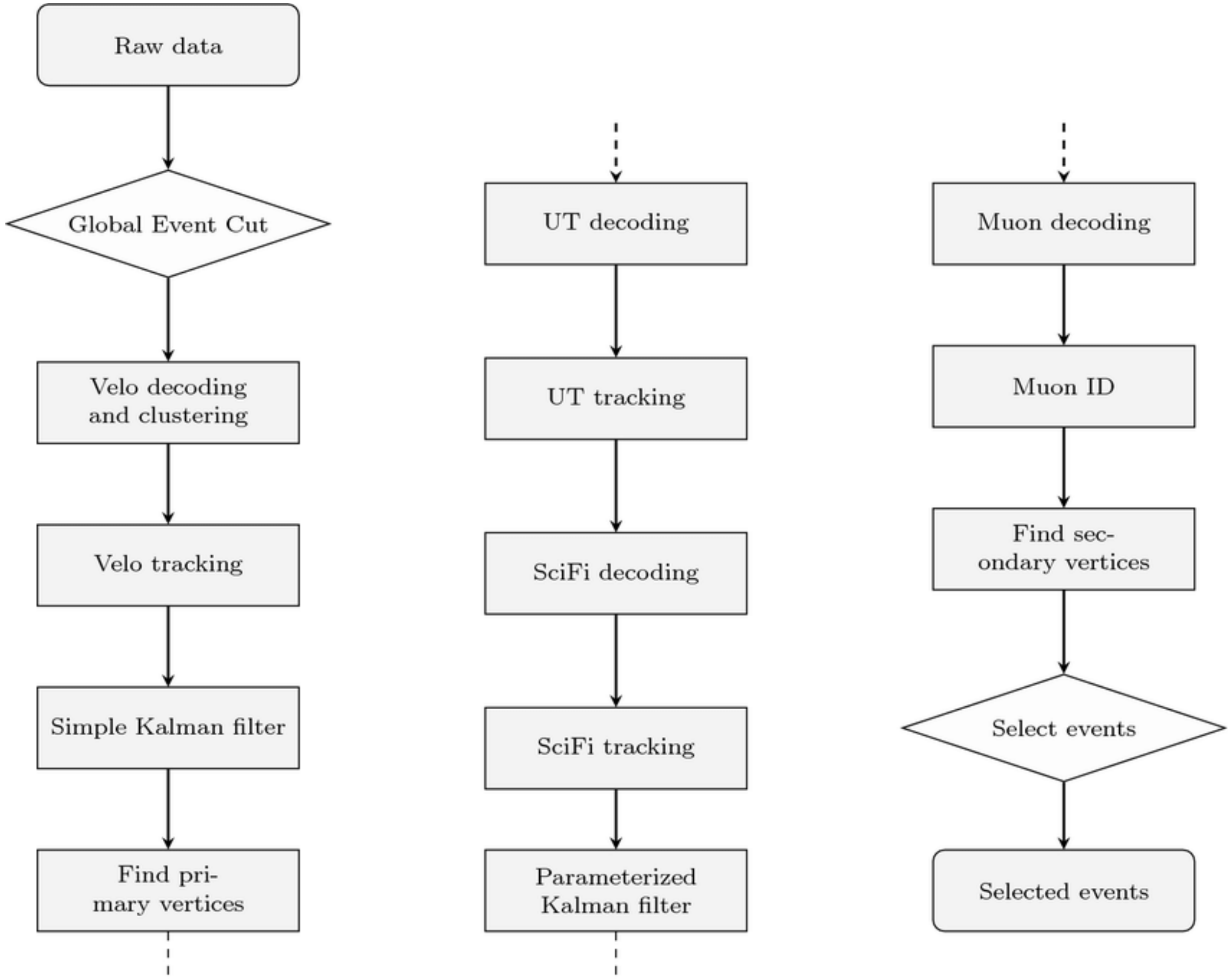
HLT1 heavily based on tracking

- HLT1 performs full upfront charged particle reconstruction down to a momentum threshold of 3 GeV/c
- For LHCb's physics purposes, different types of tracks must be reconstructed



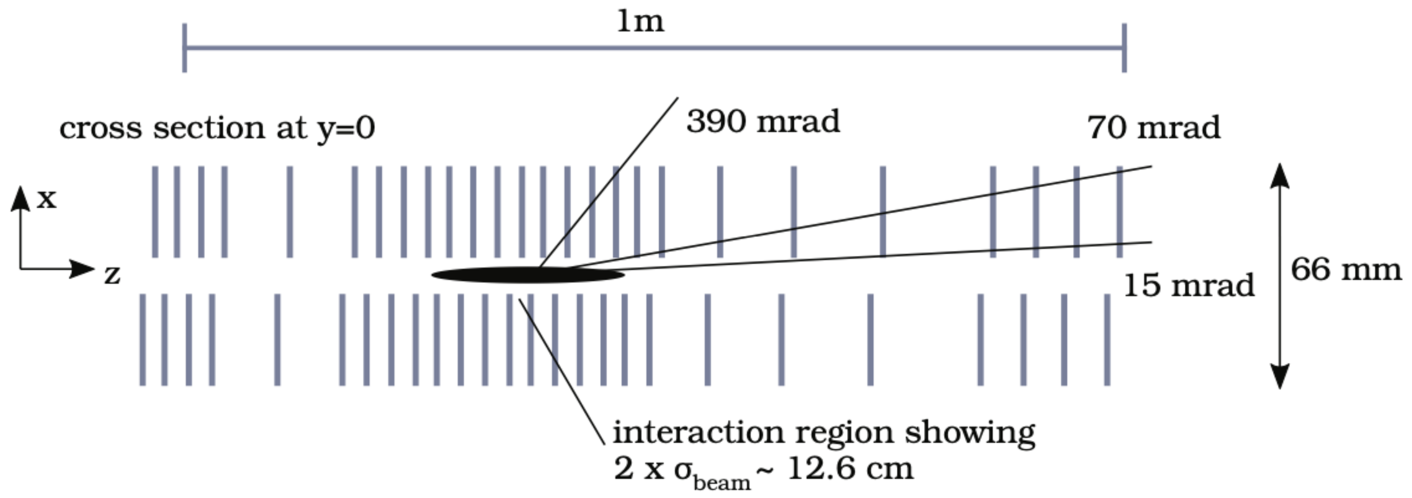
HLT1 tracking sequence

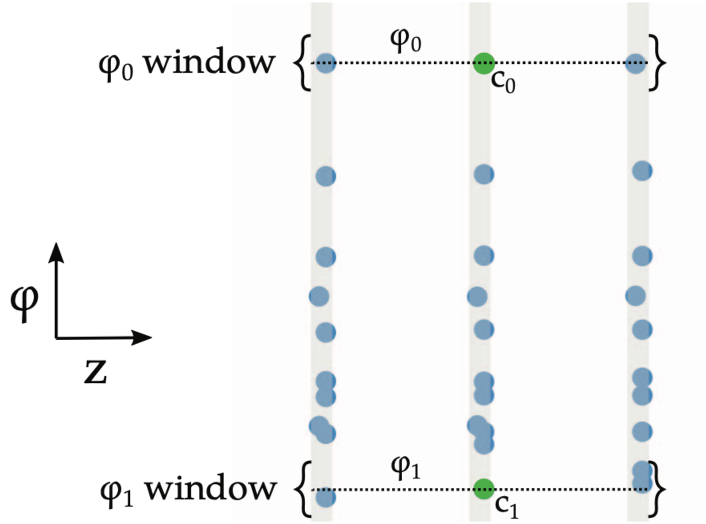
- Sequence of algorithms inside Allen project
- Reading and decoding info from tracking subdetectors
- Reconstruction:
  - Velo: tracking and vertexing
  - UT, SciFi tracking
  - Track fit and secondary vertices
  - Muon ID and Calorimeter reconstruction
- Select events and proceed in the LHCb's dataflow

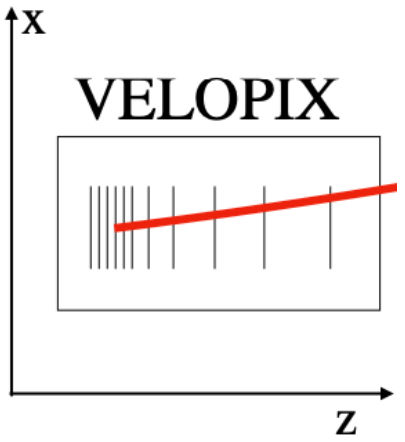


Velo tracking

- 26 silicon pixels modules providing $\sigma_{x,y} \sim 5\ \mu m$
- Straight lines trajectories
- Crucial for primary and secondary vertex finding
- Sort hits by phi and search by triplet

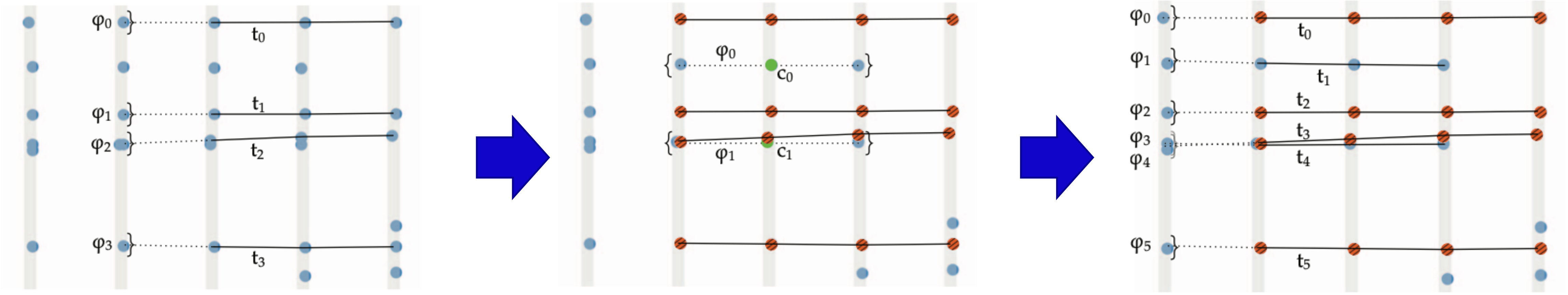






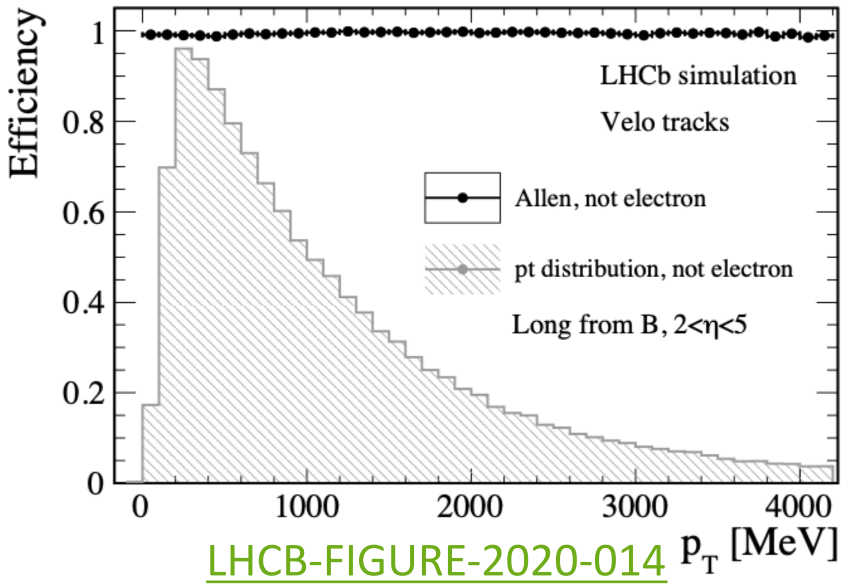
2019 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), 2019, pp. 698-707

Velo tracking



1. Creating triplets in parallel given 3 modules
2. Forward triplet to next layers
3. Repeat seeding search

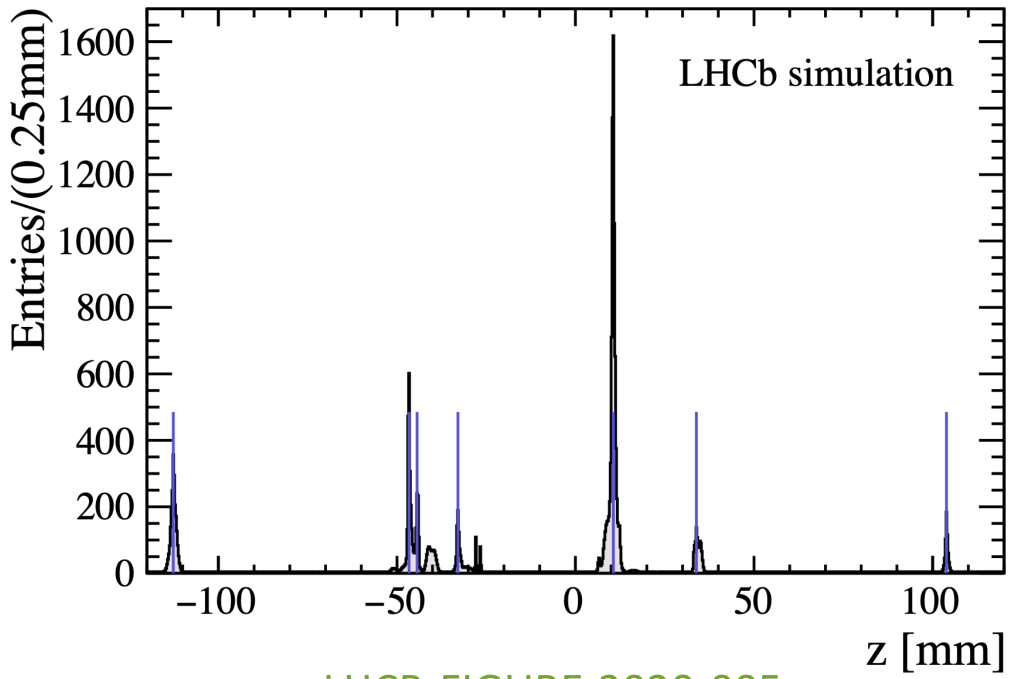
- Reconstruction efficiency of 99% for most track of interest



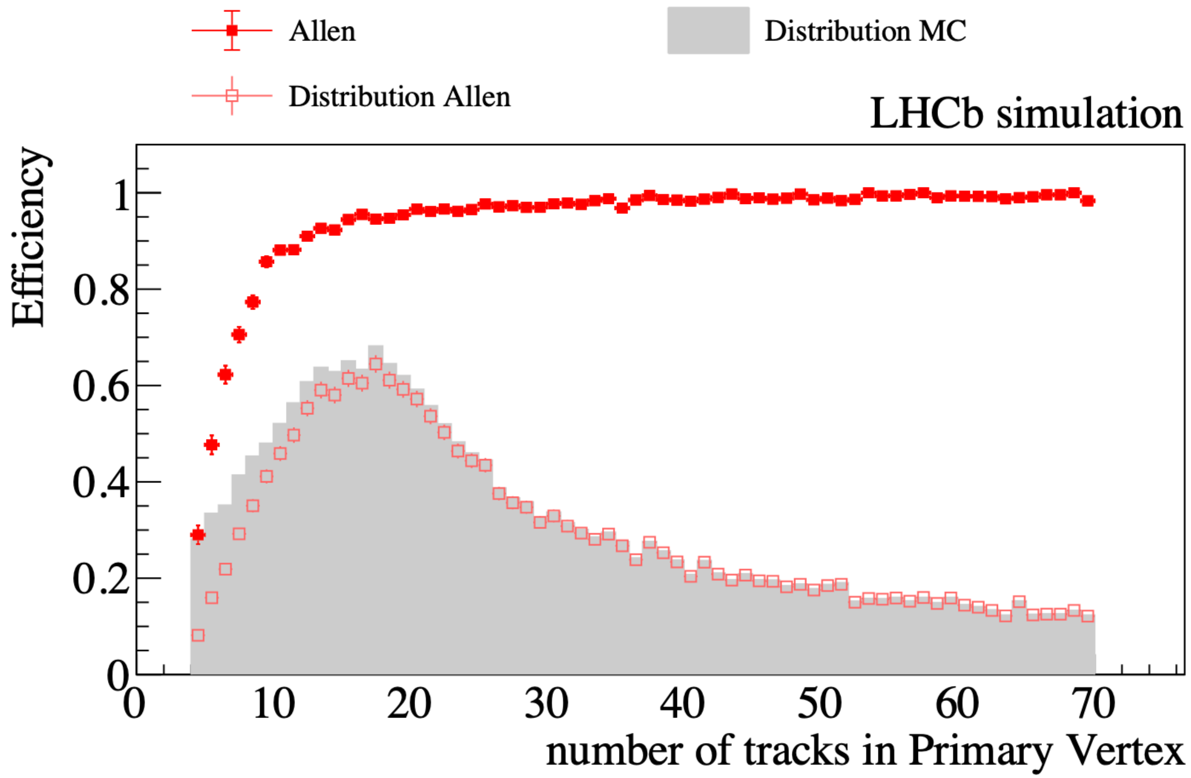
LHCB-FIGURE-2020-014

PV finding

- Extrapolate velo tracks to point of closest approach to beamline
- Cluster positions finding and fit primary vertices



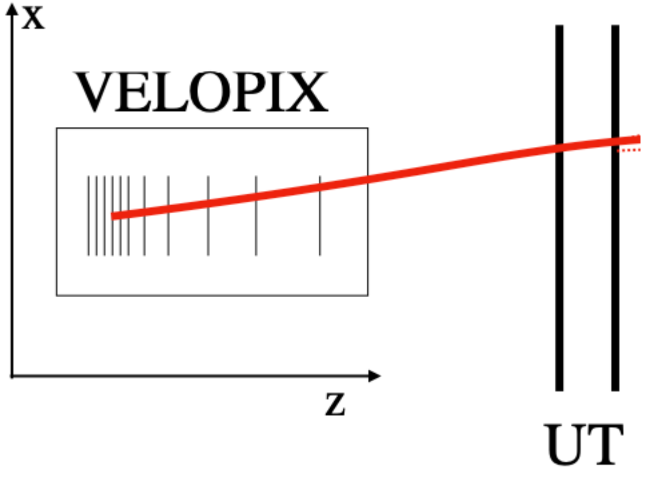
LHCB-FIGURE-2020-005

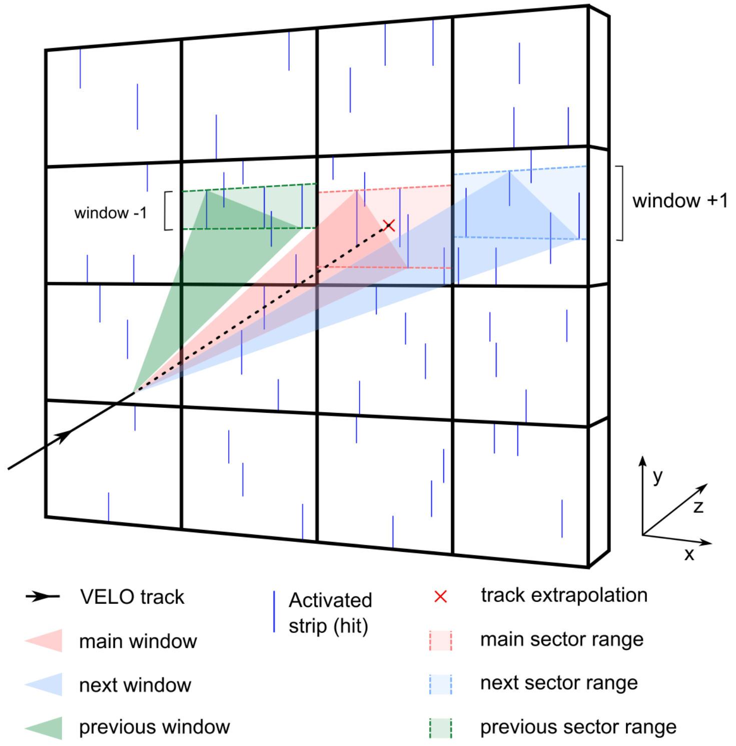


LHCB-FIGURE-2020-014

Velo-UT tracking

- 4 layers of silicon strips
- Velo tracks extrapolation taking into account small magnetic field
- Parallelize tracklets finding inside search windows requiring 3 hits

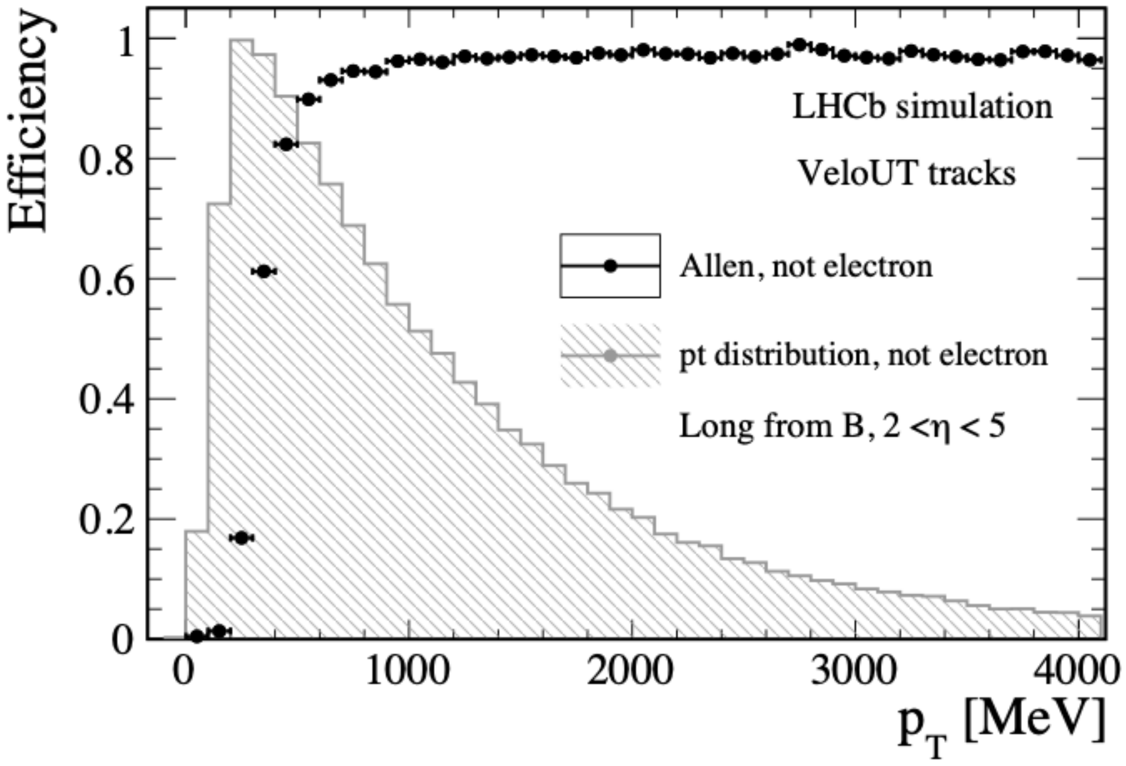




IEEE Access, vol. 7, pp. 91612-91626, 2019

Velo-UT tracking

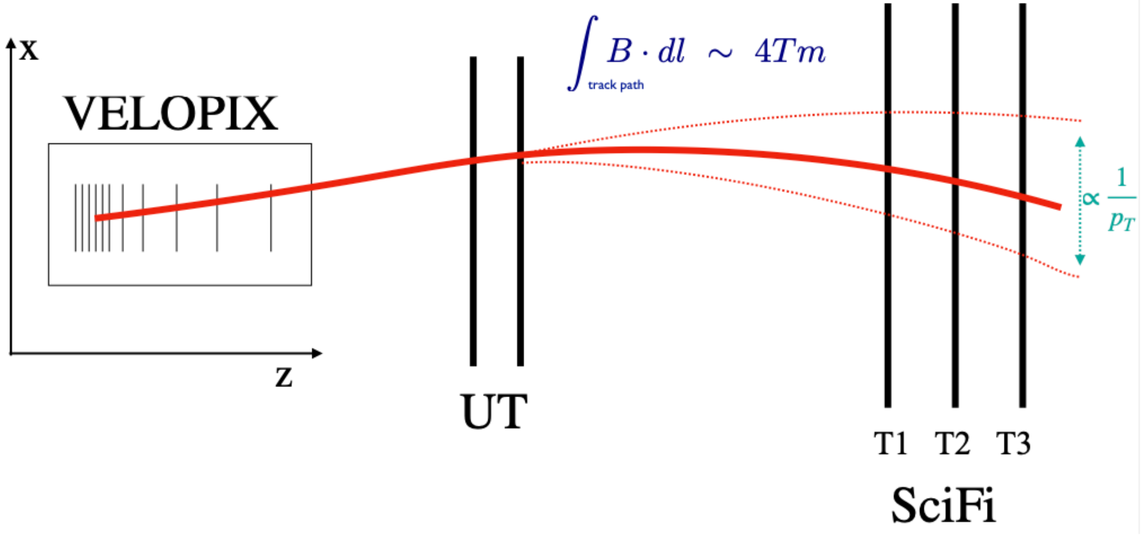
- Can reconstruct tracks:
  - With low momentum bent out of SciFi acceptance
  - For further extrapolation to the SciFi (long track)
- Provides first momentum estimate
- Helps reducing the number of combinatorics when running Forward tracking algorithm

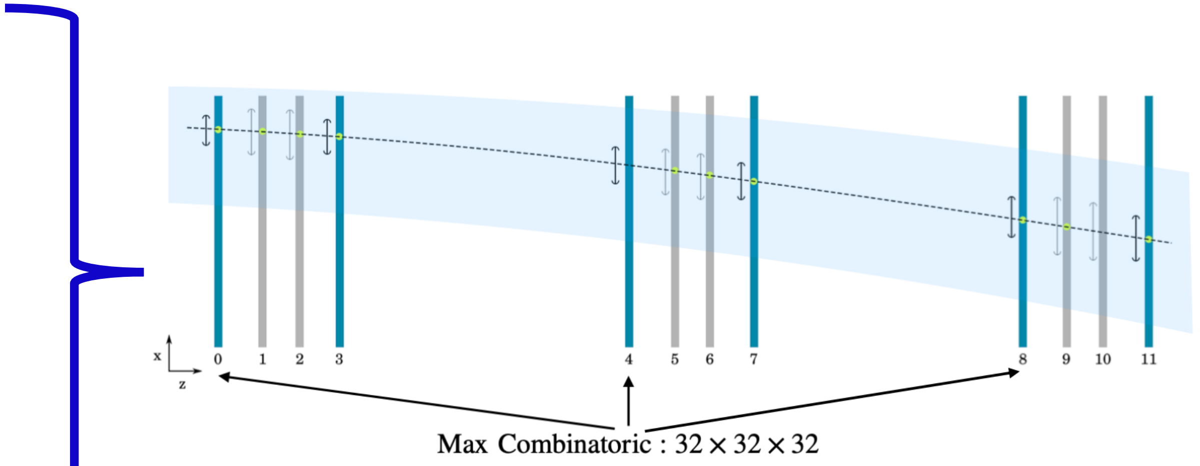


LHCB-FIGURE-2020-014

Forward tracking

- 12 layers of Scintillating Fibers (SciFi)
- Extrapolate VeloUT tracks using a B field's parametrized trajectory
- Open search windows based on the VeloUT momentum estimate
- Construct triplets and extend track to other layers (9 hits mimimum)

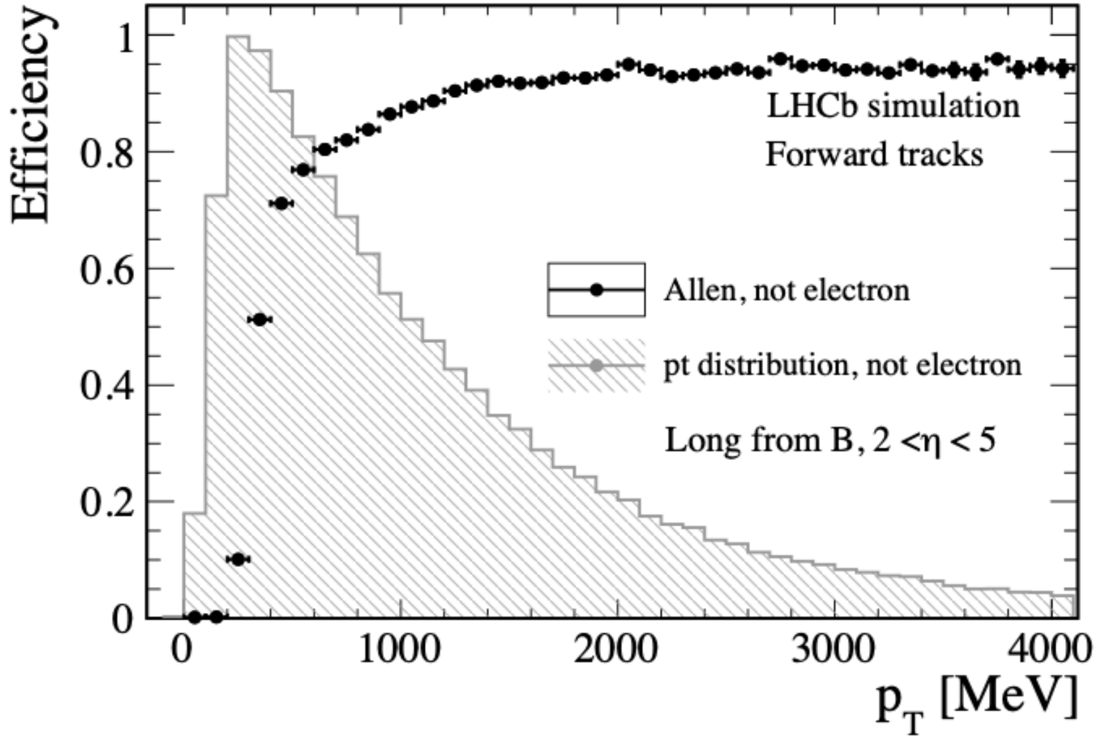


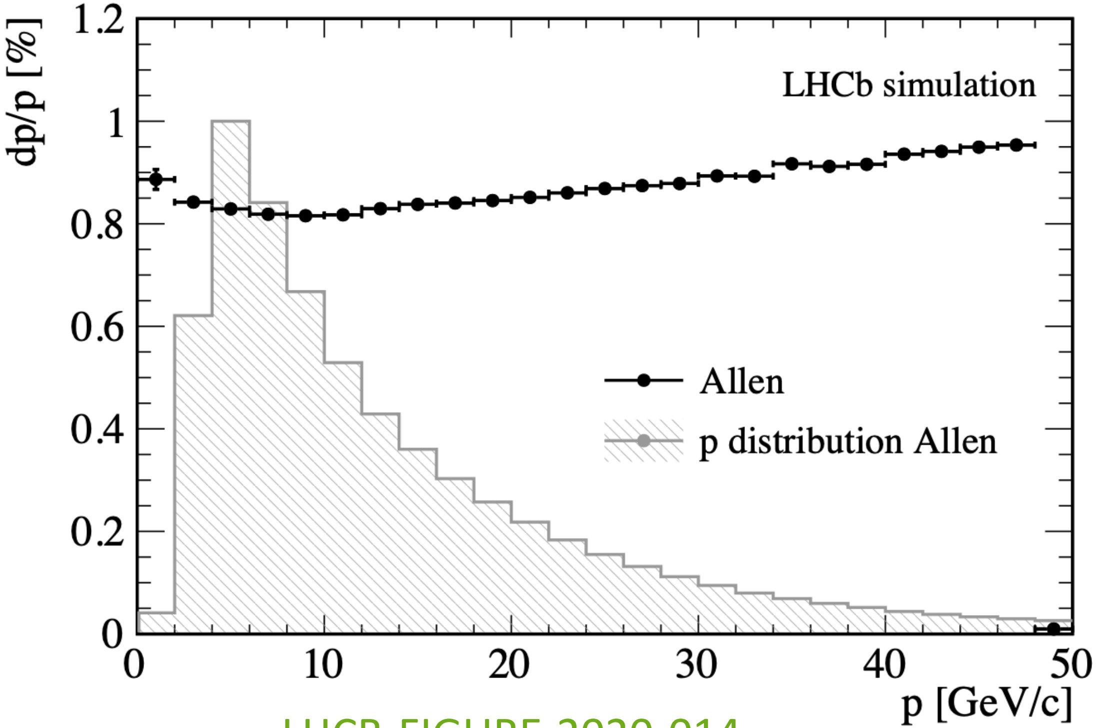


SciFi triplet finding

Forward tracking

- Reconstruct tracks with p > 3 GeV and no pt requirement
- For high momentum tracks pt > 1 GeV/c, physics efficiency plateaus ~ 90%
- Momentum resolution around 1%

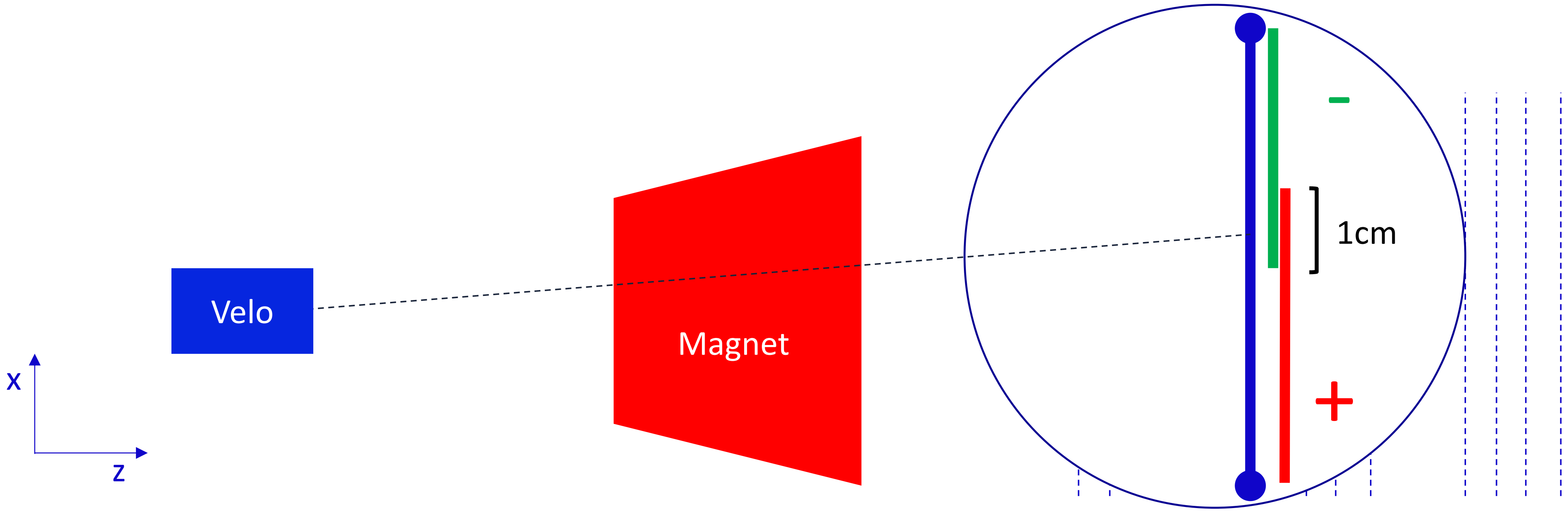




LHCB-FIGURE-2020-014

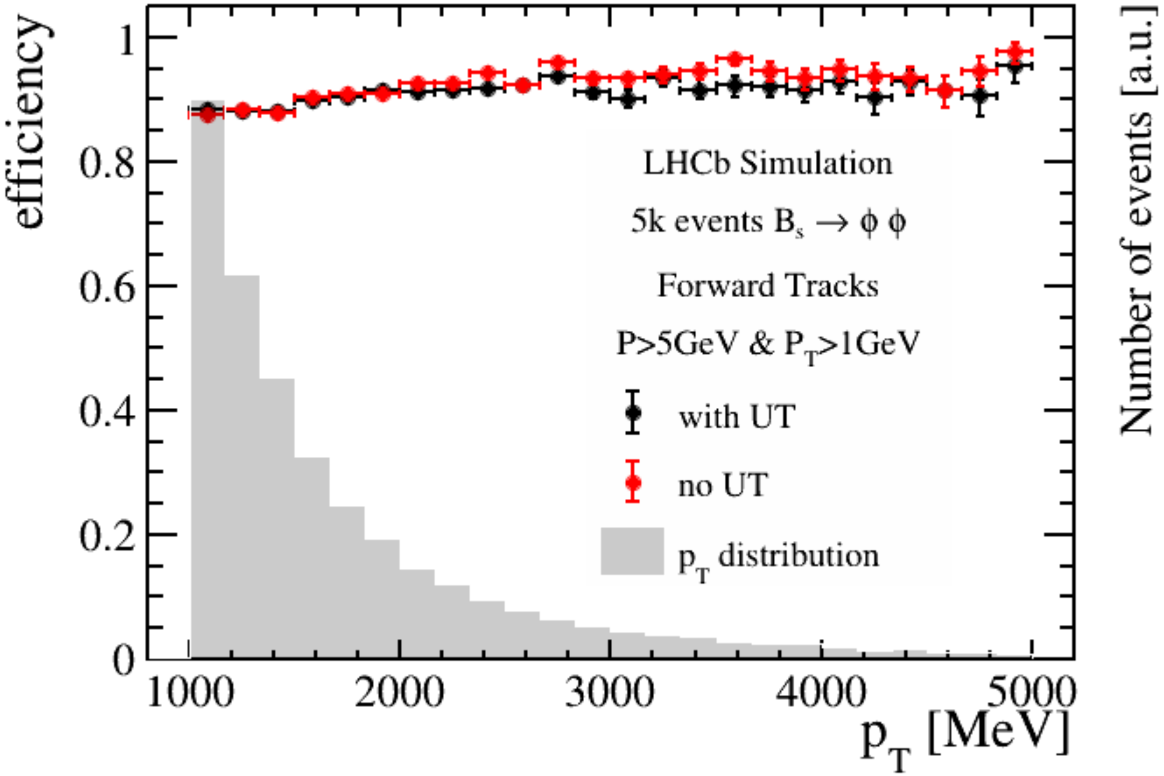
Forward tracking with no UT

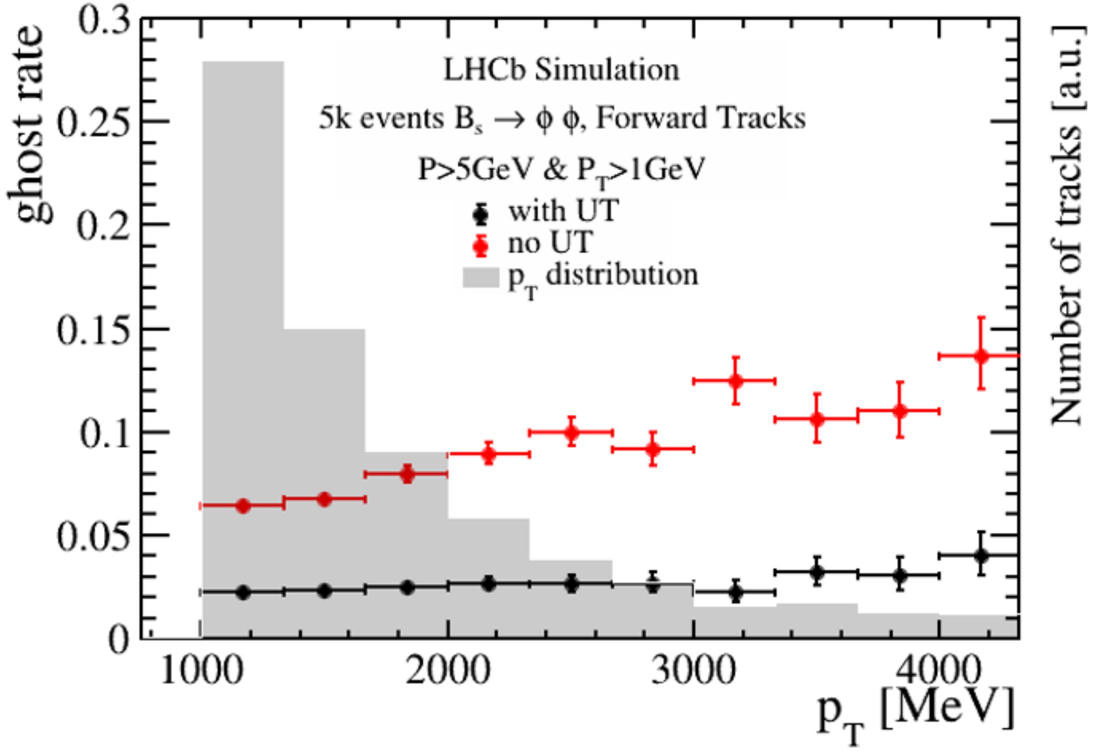
- Reconstructing tracks without info from UT
- No momentum and particle charge information
- Double sided search window strategy for high momentum tracks (p > 5 GeV and pt > 1 GeV)



Forward tracking no UT performance

- Obtaining same physics efficiency as reconstruction with UT for high momentum tracks (with an increase in ghost rate)
- Keeping the throughput comparable

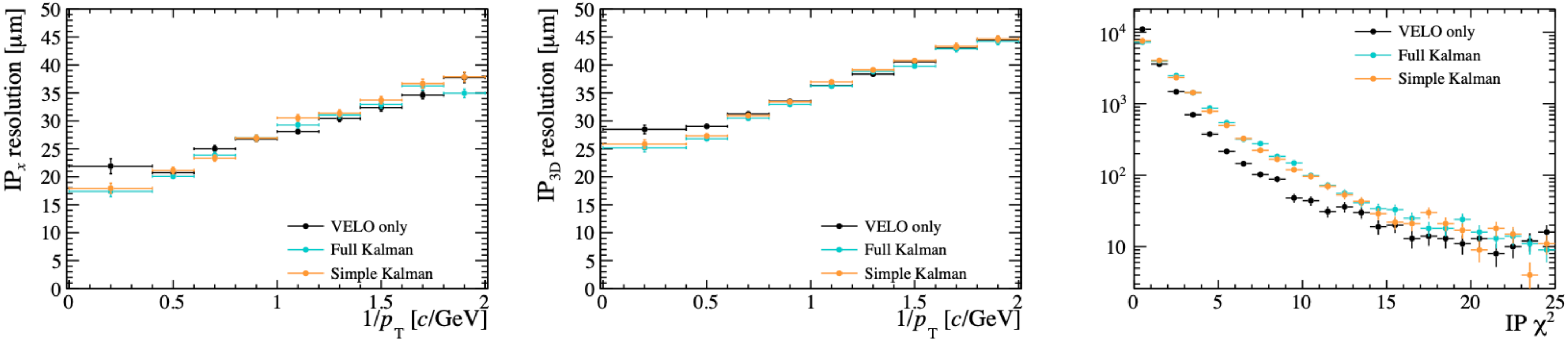




LHCb-FIGURE-2022-007

Kalman fit

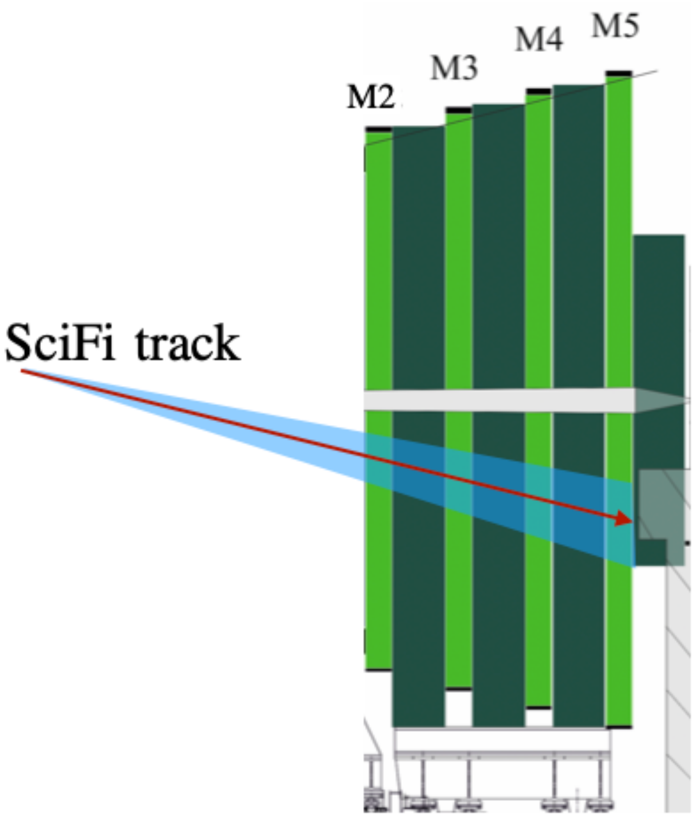
- Improves estimates of track impact parameter, momentum and helps reduce ghosts
- Different Kalman filter versions developed:
    - Full detector parametrization of the extrapolator
    - Simplified Kalman parametrization
    - Velo-only parametrization, using momentum from forward tracking

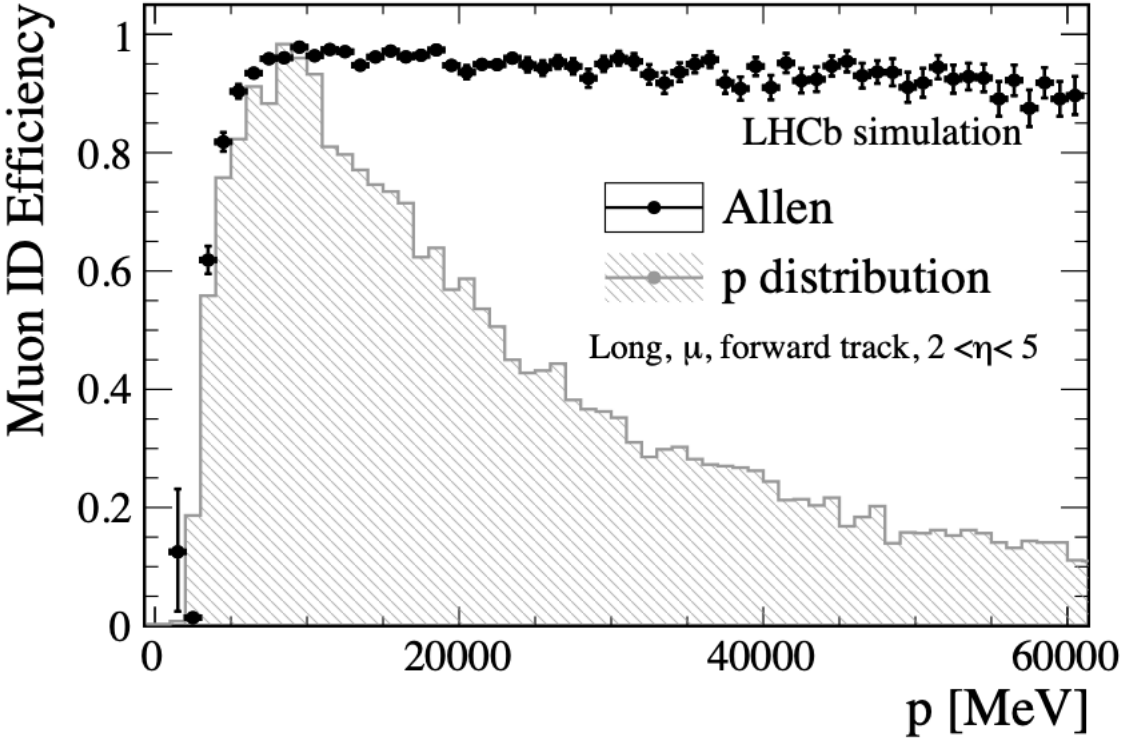


LHCb-TDR-021

Muon identification

- Forward tracks with minimum momentum 3 GeV matched with hits on Muon stations
- Very important for selection decays with muons in final state

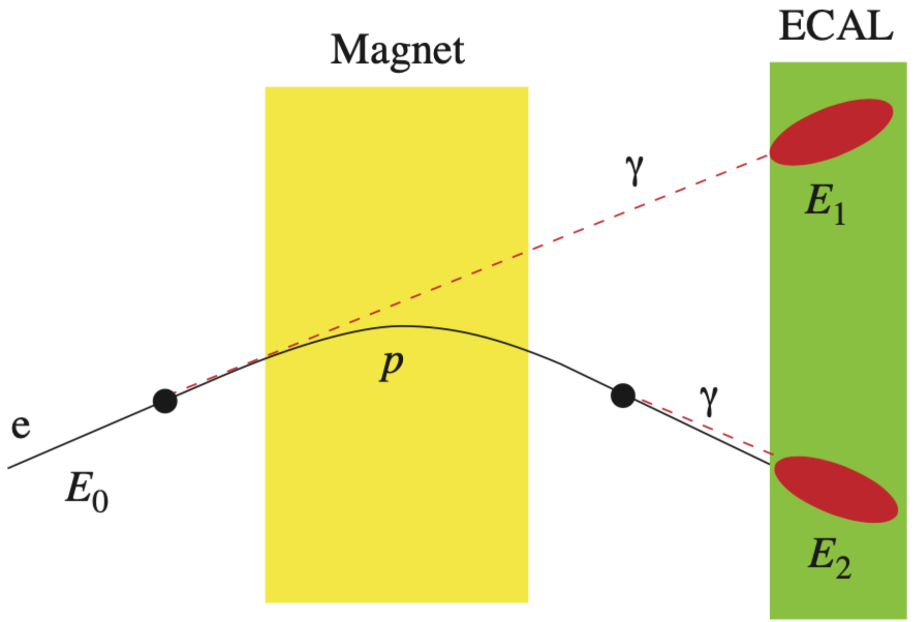




LHCB-FIGURE-2020-014

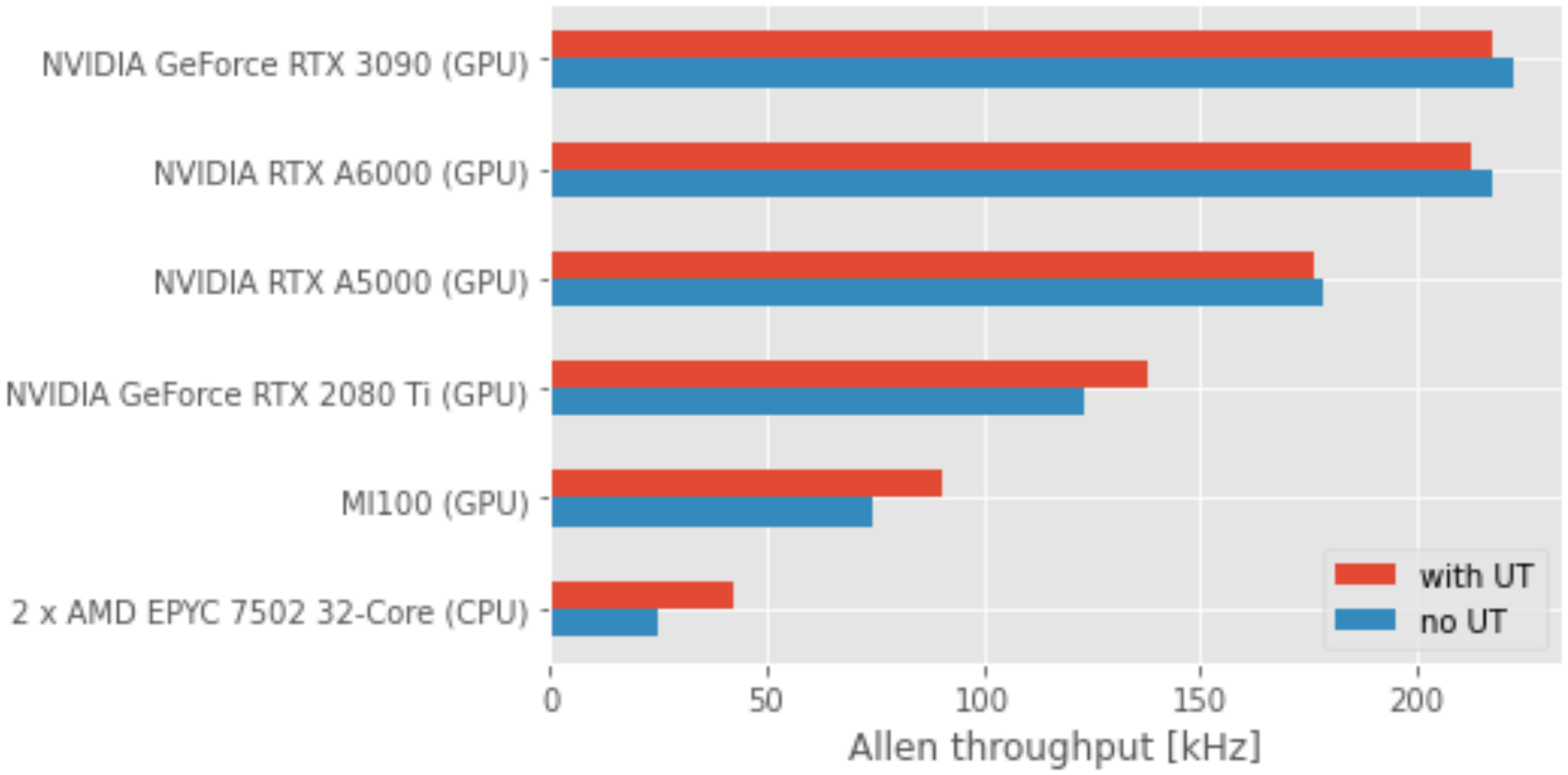
Calorimeter reconstruction

- For the first time in HLT1!
- Electron identification:
  - Extrapolating tracks to match calorimeter cells
  - Bremsstrahlung reconstruction for momentum correction
- Work ongoing for clustering/neutral reconstruction



Throughput

- HLT1 throughput measurement on different GPU cards
- RTX A5000 chosen as default
- Full input rate can be processed with ~150-200 GPUs out of 500 available slots:
  - Very promising for the future!



LHCb-FIGURE-2022-007

Selections

- Selections to fulfill broad range of decays of interest in LHCb
- Good scalability : few percent throughput loss with O(100) number of lines
- Tuning to have a combined output of 1MHz
- Selections for alignment and monitoring purposes

Trigger	Rate [kHz]
ErrorEvent	0 ± 0
PassThrough	30000 ± 0
NoBeams	5 ± 3
BeamOne	18 ± 5
BeamTwo	8 ± 3
BothBeams	4 ± 2
ODINNoBias	0 ± 0
ODINLumi	1 ± 1
GECPassthrough	27822 ± 52
VeloMicroBias	26 ± 6
TrackMVA	409 ± 23
TrackMuonMVA	23 ± 6
SingleHighPtMuon	7 ± 3
TwoTrackMVA	503 ± 26
DiMuonHighMass	131 ± 13
DiMuonLowMass	177 ± 15
DiMuonSoft	8 ± 3
D2KPi	93 ± 11
D2PiPi	34 ± 7
D2KK	76 ± 10
Total w/o pass through lines	1157 ± 39

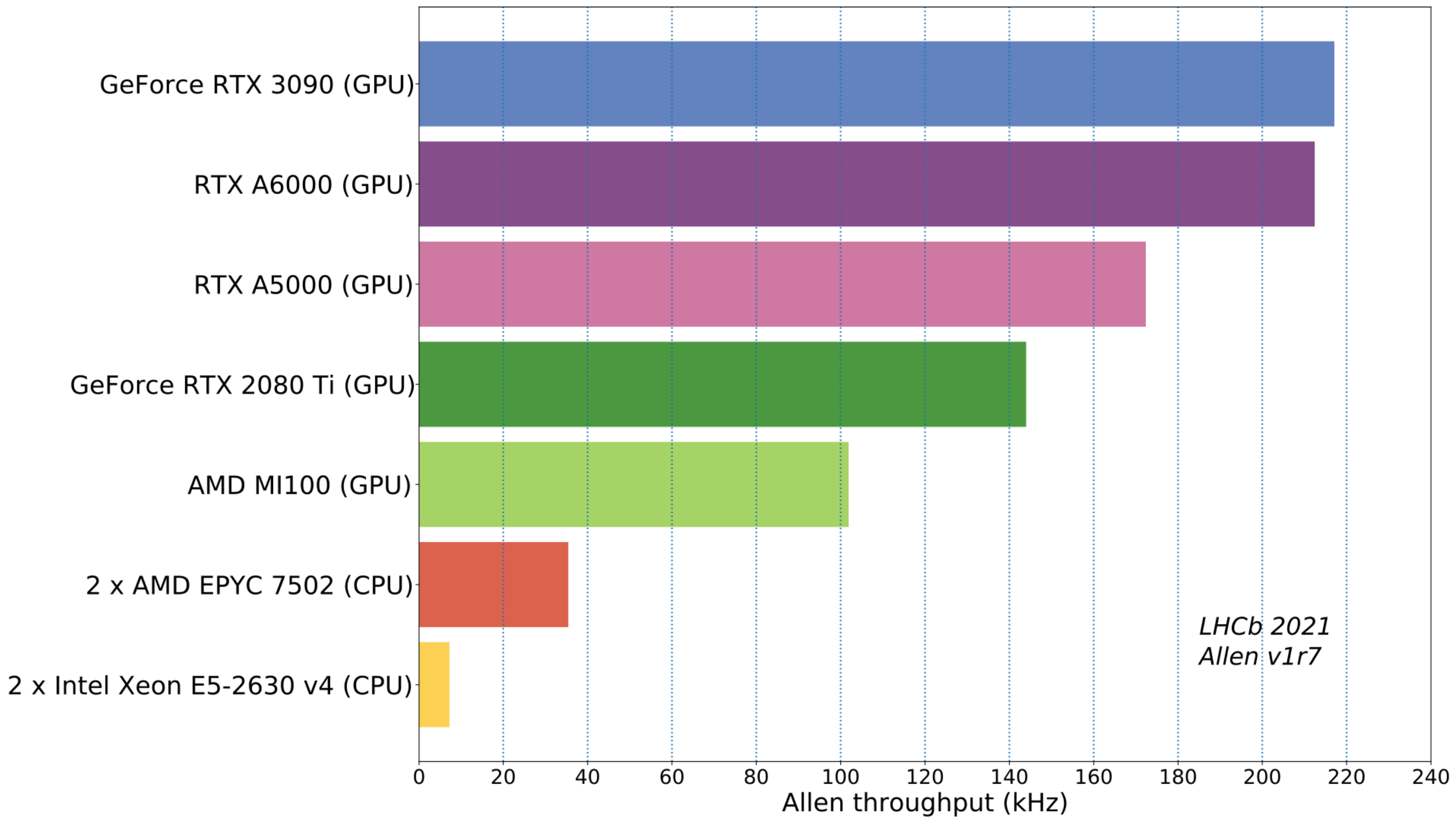
LHCB-FIGURE-2020-014

Conclusion

- LHCb is ready to face Run3:
  - 5 times higher luminosity
  - Full software trigger readout at 40 MHz
  - Looking forward to see tracks reconstructed with first collisions!
- GPUs are optimal for the First High Level Trigger tasks:
  - Tracking is highly parallelizable and very flexible for the physics use case
- Available PID information from muon stations and calorimeter
- From throughput performance, O (200) GPUs needed to take data at 40 MHz:
  - 500 available slots at LHCb
  - Higher luminosity can be handled by improved GPU technology (no saturation)
  - Very promising for the future!

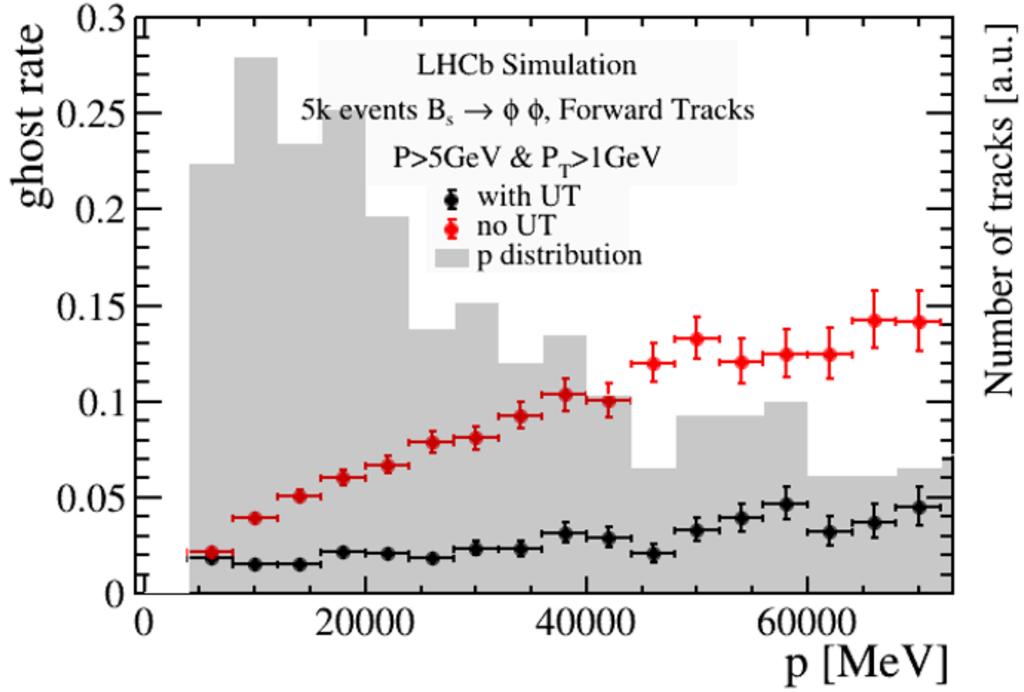
Backup

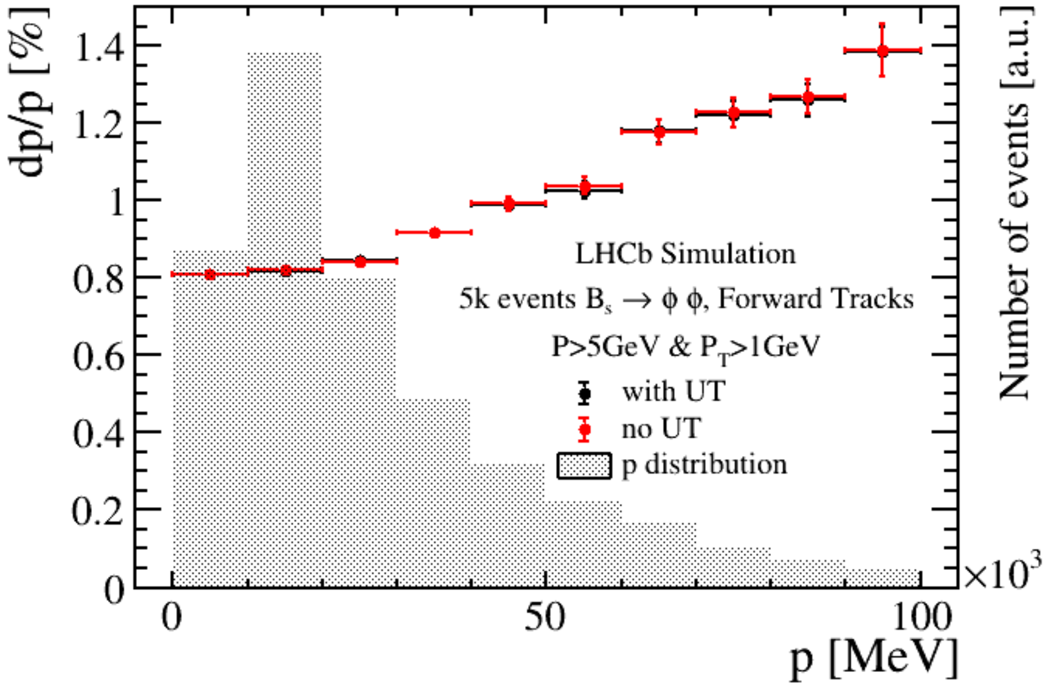
HLT1 throughput



LHCB-FIGURE-2020-014

HLT1 forward tracking





LHCb-FIGURE-2022-007

Kalman filter comparison

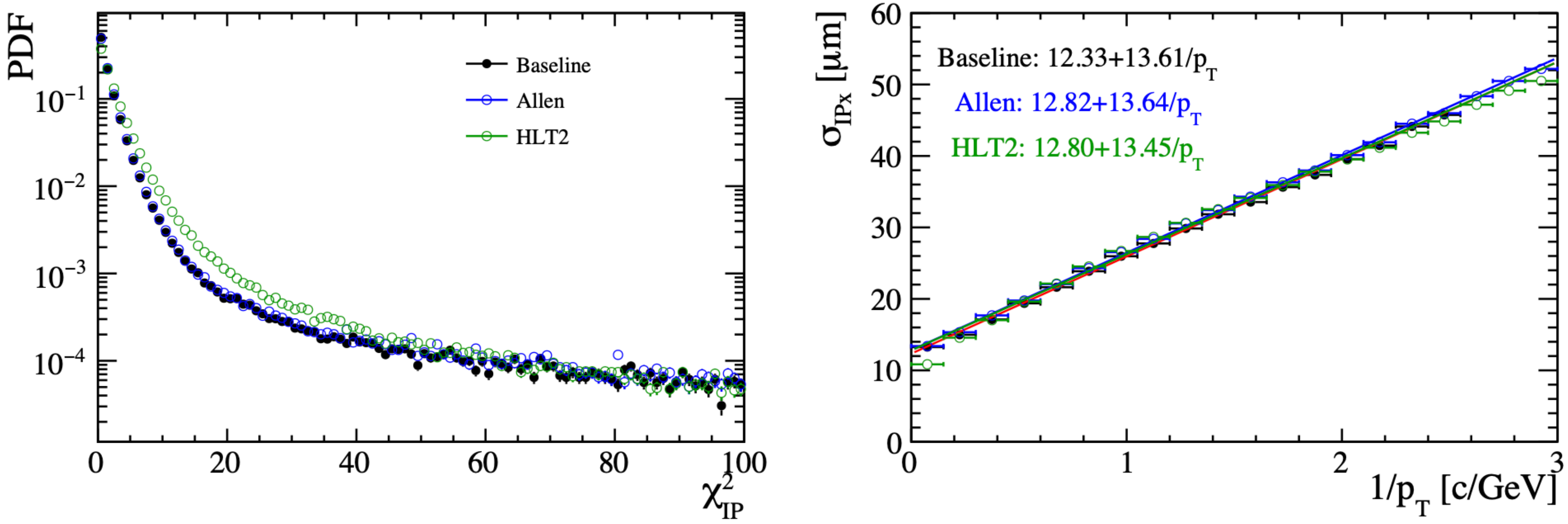


Figure 32: Comparison of (left) IP chi^2 distributions and (right) IP resolution for the CPU HLT1 Kalman filter, the Allen simplified parameterized Kalman filter, and the nominal HLT2 Kalman filter. Note that these plots were created with more restricted cuts in the UT track reconstruction algorithm: $p > 1.5$GeV and $p_T > 500$MeV.

LHCb-TDR-021